Background of the Invention

This application is a continuation-in-part of U.S. Patent Application _____ of Alizon et al. for "Cloned DNA Sequences Related to the Entire Genomic RNA of Human Immunodeficiency Virus II (HIV-2), Polypeptides Encoded by these DNA Sequences and Use of these DNA Clones and Polypeptides in Diagnostic Kits," filed January 16, 1987, which is a continuation-in-part of U.S. Patent Application Serial No. 931,866 filed November 21, 1986, which is a continuation-in-part application of U.S. Patent Application Serial No. 916,080 of Montagnier et al. for "Cloned DNA Sequences Related to the Genomic RNA of the Human Immunodeficiency Virus II (HIV-2), Polypeptides Encoded by these DNA Sequences and Use of these DNA Clones and Polypeptides in Diagnostic Kits," filed October 6, 1986 and U.S. Patent Application Serial No. 835,228 of Montagnier et al. for "New Retrovirus Capable of Causing AIDS, Antigens Obtained from this Retrovirus and Corresponding Antibodies and their Application for Diagnostic Purposes," filed March 3, 1986. The disclosures of each of these predecessor applications are expressly incorporated herein by reference.

The invention relates to cloned DNA sequences analogous to the genomic RNA of a virus known as Lymphadenopathy-Associated Virus II ("LAV-II"), a process for the preparation of these cloned DNA sequences, and their use as probes in diagnostic kits. In one embodiment, the invention relates to a cloned DNA sequence analogous to the entire genomic RNA of HIV-2 and its use as a

were detected and plaque purified. Of these phages, three were characterized by restriction mapping and South rn blot hybridization with the E2 insert and probes from its 3' end (LTR) or 5' end (envelope), as well as with HIV-1 subgenomic probes. In this instance, HIV-1 probes were used under non-stringent conditions.

A clone carrying a 9.5 kb. insert and derived from a circular viral DNA was identified as containing the complete genome and designated λROD 4. Two other clones, λROD 27 and λROD 35 were derived from integrated proviruses and found to carry an LTR and cellular flanking sequences and a portion of the viral coding sequences as shown in Figure 3, A.

Fragments of the lambda clones were subcloned into a plasmid vector p UC 18.

Plasmid pROD 27-5' is derived from λROD 27 and contains the 5' 2Kb of the HIV-2 genome and cellular flanking sequences (5' LTR and 5' viral coding sequences to the EcoRI site)

Plasmid p ROD 4-8 is dervied from λROD 4 and contains the about 5Kb HindIII fragment that is the central part of the HIV-2 genome.

Plasmid pROD 27-5' and p ROD 4.8 inserts overlap.

Plasmid pROD 4.7 contains a HindIII 1.8 Kb fragment from λROD 4. This fragment is located 3' to the fragment subcloned into pROD 4.8 and contains about 0.8 Kb of viral coding sequences and the part of the lambda phage (λL47.1) left arm located between the BamHl and HindIII cloning sites.

Plasmid pROD 35 contains all the HIV-2 coding sequences 3' to th EcoRI site, the 3' LTR and about 4 Kb of cellular flanking sequences.

Plasmid pROD 27-5' and pROD 35 in E. coli strain HB 101 are deposited respectively under No. 1-626 and 1-633 at the CNCM, and have also been deposited at the NCIB (British Collection). These plasmids are depicted in Figure 5. Plasmids pROD 4-7 and pROD 4-8 in E. coli strain TG1 are deposited respectively under No. 1-627 and 1-628 at the CNCM.

To reconstitute the complete HIV-2 ROD genome, pROD 35 is linearized with EcoRI and the EcoRI insert of pROD 27-5' is ligated in the correct orientation into this site.

The relationship of HIV-2 to other human and simian retroviruses was surmised from hybridization experiments. The relative homology of the different regions of the HIV-1 and 2 genomes was determined by hybridization of fragments of the cloned HIV-1 genome with the labelled λROD 4 expected to contain the complete HIV-2 genome (Figure 3, B). Even in very low stringency conditions (Tm-42°C.), the hybridization of HIV-1 and 2 was restricted to a fraction of their genomes, principally the gag gene (dots 1 and 2), the reverse transcriptase domain in pol (dot 3), the end of pol and the Q (or sor) genes (dot 5) and the F gene (or 3' orf) and 3' LTR (dot 11). The HIV-1 fragment used to detect the HIV-2 cDNA clones contained the dot 11 subclone, which hybridized well to HIV-2 under non-stringent conditions. Only the signal from dot 5 persisted after stringent washing. The envelope gene,

Example 4: Complete Genomic Sequence of the ROD HIV-2 Isolate

Experimental analysis of the HIV-2 ROD isolate yielded the following sequence which represents the complete genome of this HIV-2 isolate. Genes and major expression products identified within the following sequence are indicated by nucleotides numbered below:

1) GAG gene (546-2111) expresses a protein product having a molecular weight of around 55Kd and is cleaved into the following proteins:

   a) p 16 (546-950)

   b) p 26 (951-1640)

   c) p 12 (1701-2111)

2) polymerase (1829-4936)

3) Q protein (4869-5513)

4) R protein (5682-5996)

5) X protein (5344-5679)

6) Y protein (5682-5996)

7) Env protein (6147-8720)

8) F protein (8557-9324)

9) TAT gene (5845-6140 and 8307-8400) is expressed by two exons separated by introns.

10) ART protein (6071-6140 and 8307-8536) is similarly the expression product of two exons.

11) LTR:R (1-173 and 9498-9671)

env10 (1132-1215)

```
      LysGlySerAspProGluValAlaTyrMetTrpThrAsn
      AAAGGCTCAGACCCAGAAGTAGCATACATGTGGACTAAC
                .         .         .         .
CysArgGlyGluPheLeuTyrCysAsnMetThrTrpPheLeuAsn
TGCAGAGGAGAGTTTCTCTACTGCAACATGACTTGGTTCCTCAAT
          .         .      1200         .
```

env11 (1237-1305)

```
                      ArgAsnTyrAlaProCysHisIle
                      CGCAATTATGCACCGTGCCATATA
                               .         .         .
LysGlnIleIleAsnThrTrpHisLysValGlyArgAsnValTyr
AAGCAAATAATTAACACATGGCATAAGGTAGGGAGAAATGTATAT
          .         .         .      1300
```

gag1 (991-1053)

```
AspCysLysLeuValLeuLysGlyLeuGlyMetAsnProThrLeu
GACTGTAAATTAGTGCTAAAAGGACTAGGGATGAACCCTACCTTA
      1000         .         .         .
GluGluMetLeuThrAla
GAAGAGATGCTGACCGCC
     .         .
```

Of the foregoing peptides, env1, env2, env3 and gag1 are particularly contemplated for diagnostic purposes, and env4, env5, env6, env7, env8, env9, env10 and env11 are particularly contemplated as protecting agents. These peptides have been selected in part because of their sequence homology to certain of the envelope and gag protein products of other of the retroviruses in the HIV group. For vaccinating purposes, the foregoing peptides may be coupled to a carrier protein by utilizing